

(様式4)

## 学 位 論 文 の 内 容 の 要 旨

小松秀一郎 印

(学位論文のタイトル)

Deep learning-assisted literature mining for in vitro radiosensitivity data

(深層学習を活用したin vitro放射線感受性情報の解析)

(学位論文の要旨) 2,000字程度、A4判

### 【背景と目的】

がんの遺伝子変異情報に基づく治療戦略であるプレシジョンメディシンが発展を遂げているが、放射線治療の予後と関連するがん遺伝子変異プロファイルは解明されていない。これを解明するためにはがん細胞の放射線感受性と遺伝子変異のビッグデータの関連解析が有効と考えられる。遺伝子変異に関しては複数の大規模データベースが存在するが、放射線感受性に関する大規模データベースは存在しない。

培養がん細胞の放射線感受性評価の標準手法はコロニー形成法である。コロニー形成法に関して、過去の研究から (i) 2 Gy照射後の細胞生残率 (SF2) はin vitroにおいて臨床症例中に同定された放射線感受性に関連する遺伝子変異プロファイルの弁別能を有すること、(ii) SF2はメタ解析に耐えうる研究間正確性を有すること、ならびに (iii) SF2データの大多数が片対数グラフとして論文中に記載されていることが見出されている。これらのことから、既発表論文の片対数グラフからコロニー形成法SF2データを取得し、がん細胞株の放射線感受性データベースを作製すれば、放射線治療応答と関連するがん遺伝子変異プロファイル解明のための重要な生物学的基盤となると考えられる。しかし、莫大な既発表論文から網羅的にSF2データを取得するには膨大な人的・時間的コストを要する。

近年、深層学習というコンピュータテクノロジーが急速に発展している。深層学習のなかでもConvolutional neural network (CNN)と呼ばれるモデルは、画像識別コンテストにおいて2012年以降毎年優勝を続けるなど高い性能を示す。CNNは医学研究との親和性が高く、これまでに皮膚病変、眼底所見などの鑑別診断への応用可能性が示されている。

以上から、本研究はCNNを応用し既発表論文から片対数グラフとして記載されたコロニー形成法SF2を自動取得する手法の開発を目的とした。

### 【材料と方法】

2つの独立したプログラム (iSLおよびiSF) を開発した。

iSLは2つのCNNモデルおよび文字認識プログラムを組み合わせることで論文PDFから照射実験に関連する片対数グラフを自動抽出する。第一のCNNモデルであるfaster regions CNN with inception resnet v2は論文PDFからすべての図を検出し、線グラフとその他に分類するように調整した。具体的には、91分野・250万画像で予備学習済みのモデルに照射実験結果を含む既発表論文に記載された線グラフ (n = 293) およびその他の図 (n = 533) を教師データとして転移学習を施行した。第二のCNNモデルであるVGG-16は線グラフから片対数グラフを弁別するように調整した。具体的には、1000分野・120万画像で予備学習済みのモデルに照射実験

結果を含む既発表論文中記載された片対数グラフ ( $n = 517$ ) およびその他の線グラフ ( $n = 568$ ) を教師データとして転移学習を施行した。文字認識プログラムであるOCR (Google Vision API) は片対数グラフに付随するラベルから照射実験に関連性の高いキーワード (Gy、survival fraction、surviving fraction) を同定するように調整した。

iSFはCNNモデル、文字認識プログラム、および計算プログラムを組み合わせることで片対数グラフからSF2を自動抽出する。CNNモデルであるMask R-CNNは片対数グラフから生存曲線、X軸、Y軸を抽出するように調整した。具体的には、照射実験結果を含む既発表論文中に記載された片対数グラフ ( $n = 517$ ) を教師データとして転移学習を行った。文字認識プログラムであるOCRはX軸上の線量を示すラベルとY軸上の細胞生残率を示すラベルを同定し、それらの位置情報を抽出するように調整した。計算プログラムであるNumpy (Python) は生存曲線とラベルの位置情報を用いてSF2を算出するように調整した。

iSLの精度を、7種のヒト固形がん細胞株 (H460, DU145, HT-29, PC-3, H1299, HCT-116, SW480) を用いた照射実験に関連する論文 ( $n = 5606$ ) をテストデータとして検証した。iSFの精度を、コロニー形成法により取得された放射線感受性を示す片対数グラフ ( $n = 222$ ) をテストデータとして検証した。これらのテストデータを放射線腫瘍医がアノテーションし、ground truthを作製した。

## 【結果】

テストデータにおいて、iSLは感度 $91.2\% \pm 3.5\%$ 、特異度 $90.7\% \pm 3.6\%$ と高精度を示した。iSFはground truthとの誤差 $2.9\% \pm 2.6\%$ という高精度を示した。iSLおよびiSFは放射線治療専門医と比較してそれぞれ17倍、35倍の処理速度を示した。

## 【結論】

CNNを応用し既発表論文から片対数グラフとして記載されたコロニー形成法SF2を高精度に自動取得する手法を開発した。本研究成果は、がん細胞株放射線感受性データベース作製の促進を通じて、放射線治療のプレシジョンメディン化のための研究に貢献すると示唆される。